Variations on Undirected Graphical Models and their Relationships

David Heckerman*
Christopher Meek†
Thomas Richardson‡

Abstract

We consider undirected graphical models for discrete, finite variables. Lauritzen (1996) provides alternative definitions of such models and describes their relationships. We extend his analysis by considering another definition describing conditionally specified distributions. We describe the relationship of this model class with the previously described classes. In addition, we show that all definitions capture the same set of strictly positive distributions.

Keywords: graphical model, undirected graph, local Markov property, Gibbs sampler

1 Introduction

Lauritzen (1996, Ch. 3) provides alternative definitions of undirected models for discrete variables with finite state spaces. In particular, given an undirected graph $\mathcal{G}$, he defines the following families of distributions:

- $\mathcal{M}_P(\mathcal{G})$: The family of distributions satisfying the pairwise Markov property relative to $\mathcal{G}$,
- $\mathcal{M}_L(\mathcal{G})$: The family of distributions satisfying the local Markov property relative to $\mathcal{G}$,
- $\mathcal{M}_G(\mathcal{G})$: The family of distributions satisfying the global Markov property relative to $\mathcal{G}$,
- $\mathcal{M}_F(\mathcal{G})$: The family of distributions that can be written as a product of potentials over the maximal cliques in the graph
- $\mathcal{M}_E(\mathcal{G})$: The family of distributions that can be written as a limit of positive distributions in $\mathcal{M}_F(\mathcal{G})$.

In addition, he defines families of distributions limited to strictly positive distributions among the first four families just listed, denoted $\mathcal{M}^+_P(\mathcal{G})$, $\mathcal{M}^+_L(\mathcal{G})$, $\mathcal{M}^+_G(\mathcal{G})$, and $\mathcal{M}^+_F(\mathcal{G})$, respectively. He proves that these positive families are equal, and denotes the common family $\mathcal{M}^+(\mathcal{G})$. He further proves that $\mathcal{M}^+(\mathcal{G}) \subseteq \mathcal{M}_F(\mathcal{G}) \subseteq \mathcal{M}_E(\mathcal{G}) \subseteq \mathcal{M}_G(\mathcal{G}) \subseteq \mathcal{M}_L(\mathcal{G}) \subseteq \mathcal{M}_P(\mathcal{G})$ and demonstrates, with examples, that each inclusion is strict. A summary of this work is shown in the Venn diagram in Figure 1a.

In this note, we consider another definition of an undirected graphical model for $\mathcal{G}$: the family of conditionally specified distributions, which we denote $\mathcal{M}_C(\mathcal{G})$. Conditionally specified distributions have been described in Arnold, Castillo, and Sarabia (1999) and Heckerman, Chickering, Meek, Rounthwaite and Kadie (2000). We analyze the relationships between this family and those defined in Lauritzen (1996). A summary of our work is shown in Figure 1b.

2 Notation and Definitions

In this section, we review some basic graph-theoretic definitions and notation and define conditionally specified graphical models.

We use $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ to denote an undirected graph where $\mathbf{V} = \{A, B, C, \ldots\}$ denotes the set of vertices and $\mathbf{E}$ is the set of edges. We will denote an edge between two vertices A and B by $A - B$ and denote the set of neighbors of a vertex A by $\mathrm{ne}(A)$.

In order to associate a distribution with a graph we associate a variable X_A with each vertex A. We denote the set of variables corresponding to a set of vertices $\mathbf{C}$ by $\mathbf{X_C}$ and denote the set of all variables by $\mathbf{X} = \mathbf{X_V}$. We assume that each variable takes a finite set of possible values. We define the *family of conditionally specified distributions for graph* $\mathcal{G}$ as follows:

$$\mathcal{M}_C(\mathcal{G}) = \{P \mid \forall A \in V, X_A \perp\!\!\!\perp_P \mathbf{X} \setminus \{X_A\} | X_{\mathrm{ne}(A)} \text{ and } \forall \mu (\forall A\ \mu P_A = \mu) \text{ implies } \mu = P\}$$

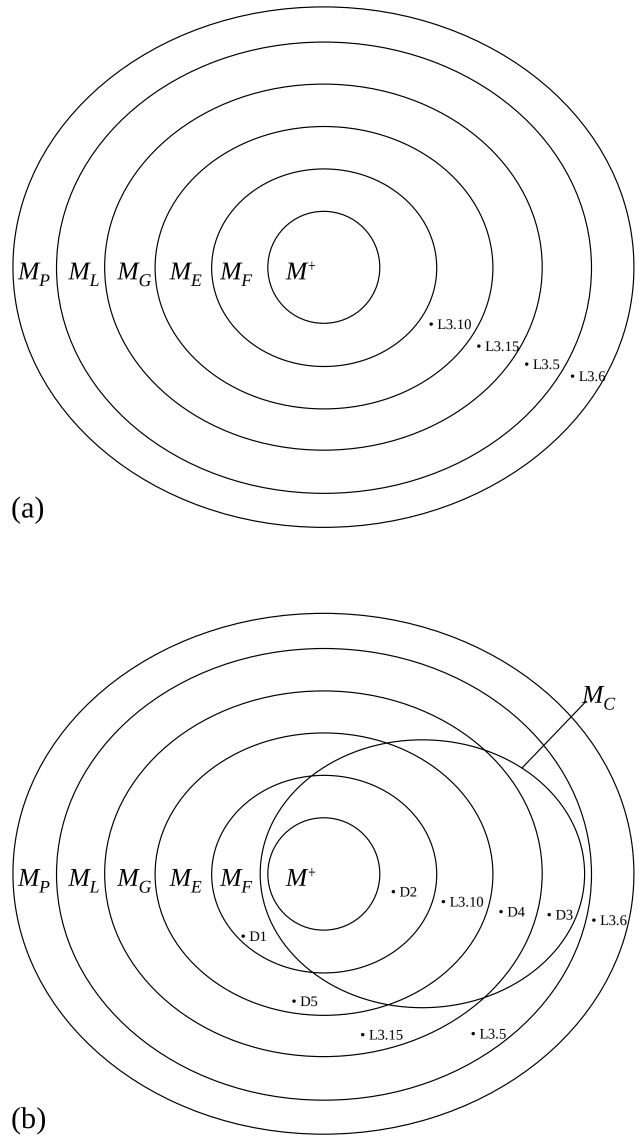

Figure 1: (a) Relationships among definitions of undirected graphical models as described by Lauritzen (1996). (b) Additional relationships among the definitions in (a) and $\mathcal{M}_C(\mathcal{G})$—the family of conditionally specified distributions. Labeled points correspond to example distributions demonstrating the non-emptiness of cells in the Venn diagram. Example distributions in Lauritzen (1996) and this note are prefixed with "L" and "D", respectively.

where $X_A \perp\!\!\!\perp_P \mathbf{X} \setminus \{X_A\} | X_{\mathrm{ne}(A)}$ means that the variable X_A is independent of the variables $\mathbf{X} \setminus \{X_A\}$ given $X_{\mathrm{ne}(A)}$ in distribution P and where $P_A = P(X_A \mid X_{\mathbf{V} \setminus \{A\}})$ is the transition matrix for a Gibbs sampler updating vertex A. In words, this is the set of distributions P which obey the local Markov property with respect to $\mathcal{G}$ (i.e., $\mathcal{M}_C(\mathcal{G}) \subseteq \mathcal{M}_L(\mathcal{G})$), and for which there exists a (single-site) Gibbs sampler whose unique stationary distribution is P. Note that, due to the restriction to local Markov distributions, $P_A = P(X_A \mid \mathbf{X} \setminus \{X_A\}) = P(X_A \mid X_{\mathrm{ne}(A)})$. Thus, a conditionally specified model is defined in terms of the family of conditional distributions whose form is defined by the graph.

In addition, let $\mathcal{M}_E^+(\mathcal{G})$ and $\mathcal{M}_C^+(\mathcal{G})$ be the families $\mathcal{M}_E(\mathcal{G})$ and $\mathcal{M}_C(\mathcal{G})$, respectively, limited to strictly positive distributions.

3 Relating the definitions

In the remainder of this note, we examine the relationship between the various families for graph $\mathcal{G}$.

3.1 Positive distributions

As mentioned, Lauritzen (1996) shows that $\mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_G^+(\mathcal{G}) = \mathcal{M}_L^+(\mathcal{G}) = \mathcal{M}_P^+(\mathcal{G})$. Here, we show that $\mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_E^+(\mathcal{G}) = \mathcal{M}_C^+(\mathcal{G})$.

Lemma 1 $\mathcal{M}_E^+(\mathcal{G}) = \mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_C^+(\mathcal{G})$.

Proof: From the definition of $\mathcal{M}_F$, we know that $\mathcal{M}_F^+(\mathcal{G}) \subseteq \mathcal{M}_E^+(\mathcal{G})$. Furthermore, we know that $\mathcal{M}_E(\mathcal{G})$ is a subset of the set of *global Markov* distributions $\mathcal{M}_G(\mathcal{G})$ (Lauritzen 1996, p. 42) and $\mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_G^+(\mathcal{G})$ (Lauritzen 1996, p. 34). From these facts we have $\mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_E^+(\mathcal{G})$.

From the definition of $\mathcal{M}_C$, we know that $\mathcal{M}_C^+(\mathcal{G}) \subseteq \mathcal{M}_L^+(\mathcal{G})$. Furthermore, the set of positive *pairwise Markov* distributions $\mathcal{M}_P^+(\mathcal{G})$ and $\mathcal{M}_G^+(\mathcal{G})$ are equal to $\mathcal{M}_L^+(\mathcal{G})$ (see, e.g., Lauritzen 1996, p. 34). From the Hammersley-Clifford Theorem, $\mathcal{M}_F^+(\mathcal{G}) = \mathcal{M}_P^+(\mathcal{G})$ (see, e.g., Lauritzen 1996, p. 36). Therefore, we have established $\mathcal{M}_C^+(\mathcal{G}) \subseteq \mathcal{M}_F^+(\mathcal{G})$. Thus, to establish the lemma we only need to show that $\mathcal{M}_F^+(\mathcal{G}) \subseteq \mathcal{M}_C^+(\mathcal{G})$. Let P be in $\mathcal{M}_F^+(\mathcal{G})$. The factorization of the distribution guarantees that P is in $\mathcal{M}_L(\mathcal{G})$. Furthermore, the conditionals P_i are all well-defined for all values of the conditioning set due to the positivity. Hence the Markov chain is irreducible and ergodic which guarantees that $P \in \mathcal{M}_C(\mathcal{G})$. QED

3.2 $\mathcal{M}_L(\mathcal{G}) \setminus (\mathcal{M}_G(\mathcal{G}) \cup \mathcal{M}_C(\mathcal{G})) \neq \emptyset$

In the remainder of Section 3, we show that certain cells in the Venn diagram of Figure 1b are not empty. We do so with specific examples. Example distributions given by Lauritzen (1996) are prefixed with the letter "L." New examples are prefixed with the letter "D." We first consider examples from Lauritzen (1996) as some of the examples we introduce are based on them.

Example 3.5 in Lauritzen (1996) illustrates that there are distributions that are locally Markov but neither globally Markov nor able to be conditionally specified.

L3.5 A distribution P for five variables X_U, X_W, X_X, X_Y, and X_Z where X_U and X_Z are independent, $P(X_U = 1) = P(X_Z = 1) = P(X_U = 0) = P(X_Z = 0) = 1/2$, and $X_W = X_U$, $X_Y = X_Z$, and $X_X = X_W X_Y$.

For the graph $\mathcal{G}_C$, U—W—X—Y—Z, Lauritzen (1996) shows that this distribution is in $\mathcal{M}_L(\mathcal{G}_C)$ but not $\mathcal{M}_G(\mathcal{G}_C)$. The distribution is not in $\mathcal{M}_C(\mathcal{G}_C)$ because any two points of support have a Hamming distance of two or more.

3.3 $\mathcal{M}_G(\mathcal{G}) \setminus (\mathcal{M}_E(\mathcal{G}) \cup \mathcal{M}_C(\mathcal{G})) \neq \emptyset$

The following distribution from Lauritzen (1996, Example 3.15) and Matus and Studeny (1995) illustrates that there are distributions that are globally Markov but neither extended Markov nor able to be conditionally specified.

L3.15 A distribution P for variables X_A, X_B, X_C, and X_D, where all variables have three possible values a, b, and c, and each of the following nine states have probability equal to 1/9:

$$\begin{array}{ccc} (a,a,a,a) & (b,a,b,c) & (c,a,c,b) \\ (a,b,b,b) & (b,b,c,a) & (c,b,a,c) \\ (a,c,c,c) & (b,c,a,b) & (c,c,b,a) \end{array}$$

For the four-cycle graph $\mathcal{G}_4$ that contains the edges $A - B$, $B - C$, $C - D$ and $A - D$, Lauritzen (1996) and Matus and Studeny (1995) show that this distribution is in $\mathcal{M}_G(\mathcal{G}_4)$ but not $\mathcal{M}_E(\mathcal{G}_4)$. The distribution is not in $\mathcal{M}_C(\mathcal{G}_4)$ because any two points of support differ in three variables.

3.4 $(\mathcal{M}_E(\mathcal{G}) \cap \mathcal{M}_C(\mathcal{G})) \setminus \mathcal{M}_F(\mathcal{G}) \neq \emptyset$

The following distribution from Lauritzen (1996, Example 3.10) and Moussouris (1974) illustrates that there are distributions that cannot be factored but are in the set of extended Markov distributions and can be conditionally specified.

L3.10 A distribution P for four binary variables X_A, X_B, X_C, X_D with support only on the points

$$\begin{array}{cccc} (0,0,0,0) & (1,0,0,0) & (1,1,0,0) & (1,1,1,0) \\ (0,0,0,1) & (0,0,1,1) & (0,1,1,1) & (1,1,1,1) \end{array}$$

and equal probability mass on each point. That is, for instance, $P(X_A = 0, X_B = 0, X_C = 0, X_D = 0) = 1/8$.

Consider the four-cycle graph $\mathcal{G}_4$, as used in the previous example. This distribution is not in $\mathcal{M}_F(\mathcal{G}_4)$ (Lauritzen 1996, p. 37) but is in $\mathcal{M}_E(\mathcal{G}_4)$ (Lauritzen 1996, p.40). It is straightforward to verify that the univariate conditionals of the distribution define a univariate Gibbs sampler with the correct stationary distribution. In particular, note that each point with support is Hamming distance one from two other points with support and that every point with support is reachable from every other point. Therefore, the distribution is in $\mathcal{M}_C(\mathcal{G}_4)$.

3.5 $\mathcal{M}_F(\mathcal{G}) \setminus \mathcal{M}_C(\mathcal{G}) \neq \emptyset$

We now come to new examples.

The following distribution illustrates that there are distributions that cannot be conditionally specified but factor.

D1 A uniform distribution for two binary random variables X_A, X_B with support only on the points $(0,0)$, $(1,1)$.

Such a distribution is in $\mathcal{M}_F(A — B)$ because the graph is complete and all distributions for two variables can be represented by the trivial factorization. The distribution is not in $\mathcal{M}_C(A — B)$ because the two points of support have a Hamming distance of two, which means that there is no way for a single-site Gibbs sampler to visit both of the points in the support of the distribution (infinitely often) without visiting points not in the support.

3.6 $(\mathcal{M}_F(\mathcal{G}) \cap \mathcal{M}_C(\mathcal{G})) \setminus \mathcal{M}_C^+(\mathcal{G}) \neq \emptyset$

The following distribution illustrates that there are non-positive distributions that factor and can be conditionally specified.

D2 A distribution for two binary random variables X_A, X_B with support only on the point (0,0).

As argued for distribution D1 above, such a distribution is in $\mathcal{M}_F(A — B)$ and $\mathcal{M}_E(A — B)$. In addition, the distribution is in $\mathcal{M}_C(A — B)$ because one can define a single-site Gibbs sampler that deterministically stays at the same point. Clearly, this distribution is not strictly positive.

3.7 $\mathcal{M}_C(\mathcal{G}) \setminus \mathcal{M}_G(\mathcal{G}) \neq \emptyset$

We now present a distribution that is in $\mathcal{M}_C(\mathcal{G})$ but not $\mathcal{M}_G(\mathcal{G})$. The construction is based on Example 3.5 in Lauritzen (1996) (see §3.2 above).

The support of the distribution in this example does not permit it to be sampled from via a univariate Gibbs sampler, since there are only four support points:

$$(0,0,0,0,0) \quad (1,1,0,0,0) \quad (0,0,0,1,1) \quad (1,1,1,1,1)$$

and they are all separated by Hamming distance at least two.

To define our distribution, we add an additional variable T to the graph, which is a neighbor of $\{U, W, X, Y, Z\}$.

D3 The distribution over 6 binary variables $\{X_T, X_U, X_W, X_X, X_Y, X_Z\}$ where $P(X_T = 1) = 0.5, P(X_U, X_W, X_X, X_Y, X_Z | X_T = 0) = 2^{-5}$ (i.e., uniform over the states) and $P(X_U, X_W, X_X, X_Y, X_Z | X_T = 1) = P(X_U | X_T = 1) P(X_Z | X_T = 1) P(X_W | X_U, X_T = 1) P(X_Y | X_Z, X_T = 1) P(X_X | X_W, X_Y, X_T = 1)$ where

$$\begin{aligned} P(X_U = 1 | X_T = 1) &= 0.5 \\ P(X_Z = 1 | X_T = 1) &= 0.5 \\ P(X_W = a | X_U = a, X_T = 1) &= 1.0 \\ P(X_Y = a | X_Z = a, X_T = 1) &= 1.0 \\ P(X_X = ab | X_W = a, X_Y = b, X_T = 1) &= 1.0 \end{aligned}$$

where a, b are the possible values (0 or 1) of the variables.

The resulting distribution $P(X_T, X_U, X_W, X_X, X_Y, X_Z)$ has support on the 36 point space:

$$\{0\} \times \{0,1\}^5 \ \cup \ \{1\} \times \{(0,0,0,0,0), (1,1,0,0,0), (0,0,0,1,1), (1,1,1,1,1)\}$$

It is simple to see that there is now a path between any two points in the support space such that each pair of adjacent points has Hamming distance one. The Markov chain resulting from the Gibbs sampler will be irreducible and ergodic, and hence will have a unique limiting distribution. This limiting distribution will be as described above.

It now only remains to observe that the distribution obeys all of the conditional independence relations required by the Local Markov property applied to the graph:

$$\begin{aligned}
X_U &\perp\!\!\!\perp X_X, X_Y, X_Z \mid X_W, X_T \\
X_W &\perp\!\!\!\perp X_Y, X_Z \mid X_U, X_X, X_T \\
X_X &\perp\!\!\!\perp X_U, X_Z \mid X_W, X_Y, X_T \\
X_Y &\perp\!\!\!\perp X_U, X_W \mid X_X, X_Z, X_T \\
X_Z &\perp\!\!\!\perp X_U, X_W, X_X \mid X_Y, X_T
\end{aligned}$$

but does not obey the global property, since

$$X_W \not\perp\!\!\!\perp X_Y \mid X_X, X_T.$$

3.8 $(\mathcal{M}_C(\mathcal{G}) \cap \mathcal{M}_G(\mathcal{G})) \setminus \mathcal{M}_E(\mathcal{G}) \neq \emptyset$

We apply a similar construction to that used in Section 3.7 to Example 3.15 in Lauritzen (1996) (see §3.3).

D4 The distribution over ternary variables X_A, X_B, X_C, X_D and binary variable X_T, where $P(X_T = 0) = P(X_T = 1) = 0.5$, $P(X_A, X_B, X_C, X_D | X_T = 0) = 3^{-4}$, and $P(X_A, X_B, X_C, X_D | X_T = 1)$ is the distribution specified in L3.15.

The resulting distribution is globally Markov with respect to the graph $\mathcal{G}$ given by:

$A — T — B \qquad C — T — D \qquad A — B — C — D — A$

because given $X_T = 0$, the variables X_A, X_B, X_C, X_D are marginally independent, whereas we have

$$X_A \perp\!\!\!\perp X_C \mid X_B, X_D, X_T = 1$$

since the conditional distribution over (X_A, X_C) is degenerate. Likewise, we have

$$X_B \perp\!\!\!\perp X_D \mid X_A, X_C, X_T = 1.$$

To demonstrate that the distribution is not in $\mathcal{M}_E(\mathcal{G})$, we follow the argument given in Lauritzen (1996), p.42. First note that each of the pairs (X_A, X_B), (X_B, X_C), (X_C, X_D), (X_A, X_D) are uniformly distributed under P given $T = 0$ and given $T = 1$. Hence, each of the triples

$$(X_A, X_B, X_T), (X_B, X_C, X_T),$$

$$(X_C, X_D, X_T), (X_A, X_D, X_T)$$

are uniformly distributed on the 18 element space: $\{a, b, c\} \times \{a, b, c\} \times \{0, 1\}$. Hence P has the same clique marginals as the uniform distribution Q defined by taking $Q(X_A, X_B, X_C, X_D, X_T) = 2^{-1}3^{-4}$. Q is clearly in $\mathcal{M}_E(\mathcal{G})$. It then follows by Lemma 3.14 in Lauritzen (1996) that $P \notin \mathcal{M}_E(\mathcal{G})$, since distributions in $\mathcal{M}_E(\mathcal{G})$ are uniquely identified via their clique marginals.

Finally, we observe that $P \in \mathcal{M}_C(\mathcal{G})$ since every point with support is reachable from every other point via a sequence of points which are Hamming distance one apart.

3.9 $\mathcal{M}_E(\mathcal{G}) \setminus (\mathcal{M}_F(\mathcal{G}) \cup \mathcal{M}_C(\mathcal{G})) \neq \emptyset$

The following distribution illustrates that there are distributions in the set of extended Markov distributions that do not factor and cannot be conditionally specified.

D5 A distribution P over 5 binary variables $\{X_A, X_B, X_C, X_D, X_E\}$ with support only on the points

$$\begin{array}{cccc}
(0,0,0,0,0) & (1,0,0,0,0) & (1,1,0,0,0) & (1,1,1,0,0) \\
(0,0,0,1,0) & (0,0,1,1,0) & (0,1,1,1,0) & (1,1,1,1,0)
\end{array}$$

$$\begin{array}{cccc}
(0,1,0,0,1) & (0,1,0,1,1) & (1,1,0,1,1) & (1,0,0,1,1) \\
(0,1,1,0,1) & (0,0,1,0,1) & (1,0,1,0,1) & (1,0,1,1,1)
\end{array}$$

and equal probability mass on each point. That is, for instance, $P(X_A = 0, X_B = 0, X_C = 0, X_D = 0, X_E = 0) = 1/16$.

We consider the graph $\mathcal{G}_5$ that contains the edges $A — B$, $B — C$, $C — D$, $A — D$, $E — A$, $E — B$, $E — C$ and $E — D$. Every point in the second set of eight support points is Hamming distance greater than or equal to two from every other point in the first set of eight points in the support thus the distribution is not in $\mathcal{M}_C(\mathcal{G}_5)$. Next we show that the distribution is not in $\mathcal{M}_E(\mathcal{G}_5)$ using the same technique as Lauritzen does for his Example 3.10 (Lauritzen 1996, p. 38). Aiming at a contradiction, we assume that the distribution factors using factors $\psi_{ABE}(\cdot), \psi_{BCE}(\cdot), \psi_{CDE}(\cdot), \psi_{ADE}(\cdot)$. Then

$$1/16 = P(0,0,0,0,0) =$$

$$\psi_{ABE}(0,0,0)\psi_{BCE}(0,0,0)\psi_{CDE}(0,0,0)\psi_{ADE}(0,0,0).$$

But also

$$0 = P(0,0,1,0,0) =$$

$$\psi_{ABE}(0,0,0)\psi_{BCE}(0,1,0)\psi_{CDE}(1,0,0)\psi_{ADE}(0,0,0).$$

From which it follows that

$$\psi_{BCE}(0,1,0)\psi_{CDE}(1,0,0) = 0.$$

Since

$$1/16 = P(0,0,1,1,0) =$$

$$\psi_{ABE}(0,0,0)\psi_{BCE}(0,1,0)\psi_{CDE}(1,1,0)\psi_{ADE}(0,1,0)$$

it must be the case that $\psi_{BCE}(0,1,0) \neq 0$. It follows that $\psi_{CDE}(1,0,0) = 0$. However, this contradicts the fact that

$$1/16 = P(1,1,1,0,0) =$$

$$\psi_{ABE}(1,1,0)\psi_{BCE}(1,1,0)\psi_{CDE}(1,0,0)\psi_{ADE}(1,0,0).$$

It remains to show that $P \in \mathcal{M}_E(\mathcal{G}_5)$. We use the same proof technique as Lauritzen (1996; p. 40) and describe a sequence of distributions $P_n \in \mathcal{M}_F(\mathcal{G}_5)$ whose limit as $n \to \infty$ is distribution **D5**:

$$P_n(a,b,c,d,e) =$$

$$\frac{n^{e(1-(ab+bc+cd-ad-b-c+1))+(1-e)(ab+bc+cd-ad-b-c+1)}}{16+16n}$$

The expression in the exponent of the numerator is one for each point in the support and zero otherwise which justifies the normalizing constant. Each distribution P_n factors according to $\mathcal{G}_5$ because each product of variables that occurs in the numerator is a subset of the cliques of the graph.

4 Discussion

We have shown that all definitions agree on positive distributions. Also, for decomposable graphs $\mathcal{G}$, Lauritzen (1996) shows that $\mathcal{M}_E(\mathcal{G}) = \mathcal{M}_F(\mathcal{G})$. Because the graphs corresponding to examples **D3** and **D4** are decomposable, however, $\mathcal{M}_C(\mathcal{G}) \neq \mathcal{M}_E(\mathcal{G})$ for this class of graphs. An interesting set of questions is whether there are non-trivial classes of graphs and/or distributions for which $\mathcal{M}_C(\mathcal{G})$ is equal to one or more of $\mathcal{M}_L(\mathcal{G})$, $\mathcal{M}_G(\mathcal{G})$, $\mathcal{M}_E(\mathcal{G})$, or $\mathcal{M}_F(\mathcal{G})$.

5 Ackowledgments

We thank Steffen Laurtizen for useful discussions.

6 Reference

Agresti, A. (1990). *Categorical Data Analysis.* Wiley & Sons, New York.

Arnold, B.C., Castillo, E., Sarabia, J. (1999). *Conditional Specification of Statistical Models.* Springer-Verlag, New York.

Heckerman, D., Chickering, D.M., Meek, C., Rounthwaite, R. and Kadie, C. (2000). Dependency networks for inference, collaborative filtering, and data visualization, in *Journal of Machine Learning Research*, **1**, pp. 49-75.

Lauritzen, S.L. (1996). *Graphical Models.* Clarendon Press, Oxford.

Moussouris, J. (1974). Gibbs and Markov random systems with constraints. *Journal of Statistical Physics*, **10**, 11–33.

Matus, R. and Studeny M. (1995). Conditional independence among four random variables I. *Combinatorics, Probability and Computing*, **4**, 269-78.